

BELLCOMM, INC.

955 L'ENFANT PLAZA NORTH, S.W. WASHINGTON, D.C. 20024

COVER SHEET FOR TECHNICAL MEMORANDUM

TITLE- Decreasing Failure Rates and Some
Related Statistical Tests

TM- 68-1033-8

DATE- November 13, 1968

FILING CASE NO(S)- 101

AUTHOR(S)- B. J. McCabe

FILING SUBJECT(S)- Reliability of Computers
(ASSIGNED BY AUTHOR(S)-

ABSTRACT

Experience has indicated that certain complex systems when first put into operation are subject to a breaking-in period. During this phase failures of the system occur at a relatively high rate, but as time wears on, this high failure rate declines steadily until a stable value is reached. This marks the beginning of what has been called the effective life of the system and it would be of obvious practical value to have a statistical method for deciding when a system has entered this period of stable performance.

This memorandum presents some axioms that describe this breaking-in phenomena and explores some of their elementary consequences. The failure rate function is described and some estimates for it presented. These are in turn used to delineate a procedure for estimating the time of the stable period's inception and the mean time between failures for this period.

(NASA-CR-100899) DECREASING FAILURE RATES
AND SOME RELATED STATISTICAL TESTS
(Bellcomm, Inc.) 16 p

N79-73141

FF No. 602(A)	(PAGES)	11/14
	(NASA CR OR TMX OR AD NUMBER)	(CODE)
	(CATEGORY)	
	AVAILABLE TO NASA OFFICES AND NASA RESEARCH CENTERS ONLY	

00/60

Unclas
11426

DISTRIBUTIONCOMPLETE MEMORANDUM TO

CORRESPONDENCE FILES:

OFFICIAL FILE COPY

plus one white copy for each
additional case referenced

TECHNICAL LIBRARY (4)

Messrs. T. A. Keegan - NASA/MA-2
J. T. McClanahan - NASA/MOR
W. E. Miller, Jr. - NASA/MOG
J. P. Nolan - NASA/MOA
L. M. Robinson - NASA/TS

Messrs. L. C. Dunseith - MSC/FS
L. R. Hamilton - MSC/FS3
H. E. Leech - MSC/FS3
L. L. Linson - MSC/FS3
J. B. MacLeod - MSC/FS5
D. H. Owen, Jr. - MSC/FC

Bellcomm

Messrs. G. R. Andersen
D. R. Anselmo
D. J. Belz
H. J. Bixhorn
H. B. Bosch
A. P. Boysen, Jr.
J. O. Cappellari
K. R. Carpenter
D. A. Chisholm
D. A. Corey
B. H. Crane
D. A. DeGraaf
J. P. Downs
W. B. Gevarter
D. R. Hagner
P. L. Havenstein
W. G. Heffron
H. A. Helm
J. J. Hibbert
E. E. Hillyard
B. T. Howard

COVER SHEET ONLY TO

D. B. James
J. E. Johnson
S. L. Levie
B. H. Liebowitz
H. S. London
J. L. Marshall
J. Z. Menard
A. C. Merritt
V. S. Mummert
J. M. Nervik
B. G. Niedfeldt
J. T. Raleigh
P. E. Reynolds
I. M. Ross
P. S. Schaenman
F. N. Schmidt
R. L. Selden
R. V. Sperry
J. W. Timko
R. L. Wagner

All Members of Department 1033
Department 1024 File

SUBJECT: Decreasing Failure Rates and Some
Related Statistical Tests - Case 101

DATE: November 13, 1968

FROM: B. J. McCabe

TM-68-1033-8

TECHNICAL MEMORANDUM

1. INTRODUCTION

Let S_k denote the time at which the k -th failure of a system occurs, where time is measured from the system's installation, and repair times are neglected. Then $S_k = X_1 + \dots + X_k$, for $k=1,2,\dots$, where X_k is the time between failures $k-1$ and k , and the X_k are assumed to be independent positive random variables.

Let $N(t)$ be the number of failures occurring in the interval $(0,t)$. Thus $N(t) = k$ if and only if $S_k < t \leq S_{k+1}$ and $N(t) \geq k$ if and only if $S_k < t$. Now let $H(t) = E(N(t))$, the mathematical expectation of $N(t)$ and set $h(t) = H'(t)$ assuming this derivative exists for all $t > 0$. N , H , and h will be called respectively, the failure function, mean failure function and failure rate.*

If the system had no breaking-in period then we might assume that the X_k were identically distributed so that S_k would be a renewal process (see [2]). In this case it is known that $\lim_{t \rightarrow \infty} H(t)/t = 1/\mu$ and that $\lim_{t \rightarrow \infty} h(t) = 1/\mu$ where $\mu = E(X_1)$. We cannot, of course, assume that S_k is a renewal process but we would like to conclude that $H(t) \sim t/\mu$ and $h(t) \rightarrow 1/\mu$ where μ is the mean value of times between failures when the system has reached its stable period. Conditions ensuring this convergence are discussed in the next section.

In section 3 we discuss methods for estimating the failure rate $h(t)$ given nothing but the times at which failures have occurred up to some time T , that is, given $N(t)$, $0 \leq t \leq T$. If we know that $h(t)$ happens to be decreasing then, as will be seen, such an estimate can be significantly improved by a smoothing operation.

*The term failure rate has another standard use which must be distinguished from ours.

Finally in cases where $h(t)$ converges decreasingly to a limit $1/\mu > 0$ we make some comments on the problem of estimating μ and other parameters of interest, for example, $t^*(\epsilon)$, the smallest value such that for all $t > t^*$, $h(t) - 1/\mu < \epsilon$.

2. CONVERGENCE OF THE MEAN FAILURE FUNCTION

In this section we will formulate conditions on the X_i so that the process S_n will conform to our intuitive idea of the breaking-in process. We will then see that these conditions imply that the mean failure function $H(t)$ is asymptotically equal to t/μ where $0 < \mu < \infty$ and μ can be looked on as the "asymptotic mean time between failures."

To begin with, we should like X_i to have a tendency to be larger than X_{i-1} since the times between failures are generally getting longer. Therefore we take as our basic assumption--

- (A) - X_i is stochastically larger than X_{i-1} , that is,
 $P(X_i > a) \geq P(X_{i-1} > a)$ for all a . If X is stochastically larger than Y we write $X \succeq Y$.

The following theorem provides some insight into this definition.

Theorem 1 Let X and Y be positive random variables with distribution functions F and G . Then,

- (i) $X \succeq Y$ if and only if $F(a) \leq G(a)$ for all a
- (ii) if $X \succeq Y$ then $E(X) \geq E(Y)$
- (iii) if $X \succeq Y$ and X and Y are independent then $P(X > Y) \geq 1/2$.

Proof (i) follows directly from the definition.

To prove (ii) we first note that

$$\begin{aligned} E(X) &= \int_0^\infty x \, dF(x) = - \int_0^\infty x \, d(1-F(x)) = - \left[x(1-F(x)) \right]_0^\infty + \int_0^\infty (1-F(x)) \, dx \\ &= \int_0^\infty (1-F(x)) \, dx \end{aligned}$$

using integration by parts at the next to last step. Note that

$$x(1-F(x)) = x \int_x^{\infty} dF(y) = \int_x^{\infty} x dF(y) \leq \int_x^{\infty} y dF(y) \rightarrow 0$$

as $x \rightarrow \infty$. Thus

$$E(X) = \int_0^{\infty} (1-F(x))dx \geq \int_0^{\infty} (1-G(x))dx = E(Y) ,$$

using the corresponding formula for Y and G and part (i).

As for (iii), first suppose G is continuous.

$$\begin{aligned} P(X>Y) &= \int_0^{\infty} P(X>y|Y=y) d G(y) \\ &= \int_0^{\infty} P(X>y) d G(y) = \int_0^{\infty} (1-F(y)) d G(y) \\ &\geq \int_0^{\infty} (1-G(y)) d G(y) = 1 - \int_0^{\infty} G(y) d G(y) \\ &= 1 - G(y)G(y) \Big|_0^{\infty} + \int_0^{\infty} G(y) dG(y) = \int_0^{\infty} G(y) d G(y) , \end{aligned}$$

once again using integration by parts. But if

$$1 - \int_0^{\infty} G(y) d G(y) = \int_0^{\infty} G(y) d G(y)$$

then the latter quantity equals 1/2, so that

$$P(X > Y) \geq \int_0^{\infty} G(y) dG(y) = 1/2 .$$

An approximation argument gives the same result when G is not assumed to be continuous.

The second condition which must be placed on the distribution of the X_i is that their growth be bounded in such a way that the failure process eventually stabilizes, or converges, in some sense, to a random variable X whose distribution characterizes the way failures occur during the most effective phase of the system's operation. Thus we make the assumption--

(B) - There exists a random variable Y with finite mean α such that $X_i \lesssim Y$ for all i .

Without some assumption like (B) the times between failures might go on increasing without bound, which is certainly not the case in the systems we are interested in studying. Condition (B) is somewhat stronger than a cursory glance might indicate, if (A) is also assumed to hold, as the following theorem shows.

Theorem 2 Let X_i be a sequence of random variables satisfying (A) and (B). Then there exists a random variable X with finite mean μ such that $X_i \lesssim X$, Y_i converges in distribution to X , and $E(X_i) \rightarrow \mu$.

Proof Let F_i be the distribution of X_i , let Y be as in (B) and let G be the distribution of Y . Then the characteristic function ϕ_n of X_n is--

$$\begin{aligned} \phi_n(t) &= \int_0^{\infty} e^{itx} dF_n(x) \\ &= - \int_0^{\infty} e^{itx} d(1-F_n(x)) \\ &= - \left[e^{itx}(1-F_n(x)) \right]_0^{\infty} + \int_0^{\infty} (1-F_n(x))ite^{itx} dx \\ &= 1 + it \int_0^{\infty} (1-F_n(x)) e^{itx} dx . \end{aligned}$$

The sequence $1-F_n(x)$ is increasing and bounded above by $1-G(x)$. Therefore $\lim_{n \rightarrow \infty} 1-F_n(x) = 1-F(x)$ exists for all x .

We show that $F(x)$ is a distribution function by Levy's Continuity Theorem (see [4, p. 191]), that is, we show that $\phi_n(t)$ converges to a function which is continuous at $t=0$.

$$\begin{aligned} \lim_{n \rightarrow \infty} \phi_n(t) &= \lim_{n \rightarrow \infty} (1+it \int_0^{\infty} (1-F_n(x)) e^{itx} dx) \\ &= 1 + it \int_0^{\infty} (1-F(x)) e^{itx} dx, \end{aligned}$$

and the limit exists. The interchange of limit and integration is justified above by Lebesgue's Dominated Convergence Theorem and the fact that $|(1-F_n(x))e^{itx}| = 1-F_n(x) \leq 1-G(x)$, the latter function being integrable. We now must show that

$1 + it \int_0^{\infty} (1-F(x)) e^{itx} dx$ is a continuous function of t , at $t=0$.

But the integral is obviously bounded in a neighborhood of $t=0$, so that the whole function is continuous at 0. Thus X_n converges in distribution to X , and the other assertions of the theorem are obvious consequences of the definitions of X and F .

We are now in a position to show that if the sequence $\{X_n\}$ satisfies (A) and (B) then $H(t) \sim t/\mu$. To do this we need the following result due to Kawata [3] and Smith [7].

Theorem 3 Let $\{X_n\}$ be a sequence of non-negative independent random variables such that

$$(i) \quad \frac{1}{N} \sum_{n=1}^N E(X_n) \rightarrow \mu, \quad 0 < \mu < \infty$$

(ii) for every $\epsilon > 0$,

$$\int_{n\epsilon}^{\infty} \frac{1}{n} \sum_{k=1}^n (1-F_k(x)) dx \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Then $\lim_{t \rightarrow \infty} H(t)/t = 1/\mu$.

Now we may state the main result of this section.

Theorem 4 Let $\{X_n\}$ be a sequence of independent non-negative random variables satisfying (A) and (B). Then there exists a random variable X with mean μ such that

(i) X_n converges in distribution to X

(ii) $\lim_{t \rightarrow \infty} H(t)/t = 1/\mu$.

Proof (i) has already been established. Let $\mu_n = E(X_n)$. Then $\mu_n \rightarrow \mu$, also by Theorem 2. But this implies that also

$\frac{1}{n} \sum_{k=1}^n \mu_k \rightarrow \mu$. Let F_k and F be the distribution functions of

X_k and X . Then,

$$\int_{n\epsilon}^{\infty} \frac{1}{n} \sum_{k=1}^n (1-F_k(x)) dx \leq \frac{1}{n} \sum_{k=1}^n \int_{n\epsilon}^{\infty} (1-F(x)) dx \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Thus both conditions of Theorem 3 are satisfied and hence the theorem is proven.

3. ESTIMATION OF THE FAILURE RATE

As above it is assumed that the times between failures, X_n , are non-negative independent random variables, stochastically increasing, and converging in distribution to a random variable X with finite mean μ . Then it was shown in section 2 that

$$(*) \quad H(t) = E(N(t)) \sim t/\mu$$

where $N(t)$ is the number of failures occurring in $(0, t)$. Clearly $(*)$ is not of itself sufficient⁺ to conclude that the failure rate $h(t) = H'(t) \rightarrow 1/\mu$ as $t \rightarrow \infty$, but in the context of other renewal theorems it seems to be only a slight additional restriction to assume that, in fact, $h(t)$ converges to $1/\mu$.

Moreover, in the type of reliability problem being treated here, there is strong empirical evidence to suggest that $h(t)$ is a decreasing function. The conditions (A) and (B) of section 2 are certainly not enough to guarantee this so the following additional assumption is introduced--

- (C) - The failure rate $h(t)$ is a decreasing function such that $h(t) \rightarrow 1/\mu$ as $t \rightarrow \infty$.

A host of statistical problems concerning h now pose themselves: given $N(t)$ for $0 \leq t \leq T$, estimate

- (i) $h(t)$ for $0 \leq t \leq T$
- (ii) $h(T)$
- (iii) $t^*(\epsilon)$, the smallest number such that for all $t > t^*$, $h(t) - 1/\mu < \epsilon$.

The practical value of estimates for these quantities is quite obvious. The estimators to be presented here are of an ad hoc nature, however, and nothing is known of their moments or distributions, asymptotically or otherwise. This is a result of the great generality of the problem as posed here; on the other hand, if some stringent restrictions are placed on the distributions of the X_n some more detailed knowledge of the estimators might be had but this kind of study will be postponed for now.

3.1 A Class of Estimators for $h(t)$

We wish to estimate the function $h(t) = H'(t)$ where $H(t) = E(N(t))$, given $N(t)$ for $0 \leq t \leq T$. $N(t)$ is an unbiased estimate for $H(t)$ and so we might attempt differentiating $N(t)$ to get an estimate for h . Unfortunately $N(t)$ is a step function

⁺Consider, e.g., the function $H(t) = t + \cos t$.

and its derivative equals 0 everywhere it exists so that this procedure is not very promising. We may however borrow a technique from time series analysis for getting around this difficulty (see especially [6] for a discussion of the way in which spectral density estimation applies to probability density estimation).

Let K be a function satisfying

$$K \geq 0, K(-x) = K(x), \int_{-\infty}^{\infty} K(x) dx = 1$$

and $\lim_{x \rightarrow \infty} x K(x) = 0$. Then we set

$$\hat{h}_K(t) = \int_0^{\infty} K(x-t) dN(x) = (K*N)(t)$$

and take \hat{h}_K as an estimate of h . K is called the kernel of this transformation, and \hat{h}_K may be called the convolution of K with N . The following example will illustrate how the convolution $K*N$ provides an estimate of H' (See [6] for a proof that $K*N$ converges to H' , if N is a sample distribution function, as the size of the sample n increases and K is allowed to depend on n .)

Take $\Delta > 0$ and let

$$\begin{aligned} K(x) &= 1/\Delta, \quad |x| \leq \Delta/2 \\ &= 0, \quad |x| > \Delta/2. \end{aligned}$$

$$\begin{aligned}
\text{Then } \hat{h}_K(t) &= \int_0^{\infty} K(x-t) dN(x) \\
&= \sum_{k=1}^{\infty} K(S_k - t) \{N(S_k) - N(S_k^-)\} \\
&= \sum_{k=1}^{\infty} K(S_k - t) \\
&= \frac{1}{\Delta} \cdot (\text{number of terms } S_k \text{ within } \Delta/2 \text{ of } t) \\
&= \frac{1}{\Delta} \cdot (N(t+\Delta/2) - N(t-\Delta/2)) \\
&= \frac{1}{\Delta} \cdot (N(t+\Delta/2) - N(t)) + \frac{1}{\Delta} \cdot (N(t) - N(t-\Delta/2))
\end{aligned}$$

Thus

$$E\hat{h}_k(t) = \frac{1}{\Delta} \cdot (H(t+\Delta/2) - H(t)) + \frac{1}{\Delta} \cdot (H(t) - H(t-\Delta/2))$$

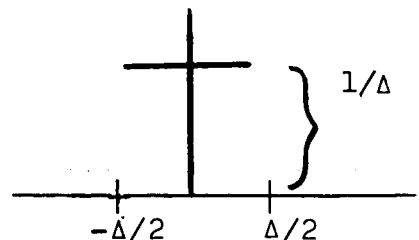
and

$$E\hat{h}_k(t) \doteq H'(t)$$

The usual procedure in problems of this sort is to perform several different transformations using a variety of kernels K since some are more sensitive to certain kinds of tremors and wiggles than others and so it is best to try a number of them. Here are three examples and the literature of time series analysis abounds with others.

Example 1

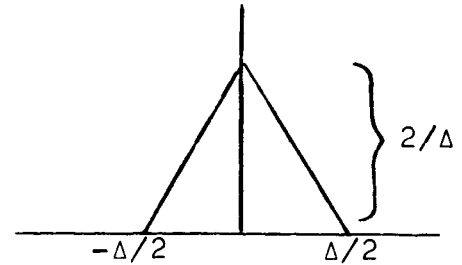
$$\begin{aligned}
K_1(x) &= 1/\Delta \text{ for } |x| \leq \Delta/2 \\
&= 0 \text{ for } |x| > \Delta/2
\end{aligned}$$



Example 2

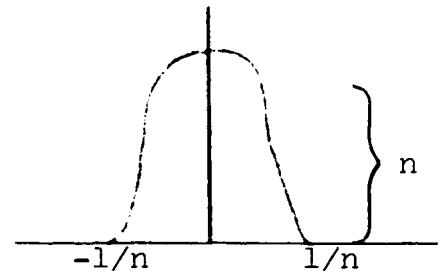
$$K_2(x) = \frac{2}{\Delta}(1 - |2x/\Delta|), \quad |x| \leq \Delta/2$$

$$= 0, \quad |x| > \Delta/2$$

Example 3

$$K_3(x) = \frac{n}{2}(1 + \cos nx), \quad |x| \leq 1/n$$

$$= 0, \quad |x| > 1/n$$



Note that each K_1 depends on a parameter Δ or n , which may be varied so as to emphasize local behavior by taking $\Delta > 0$ small or n large.

Since h is assumed to be a decreasing function it is entirely reasonable to restrict attention to those estimates of h which are themselves decreasing. This can be accomplished by smoothing any of the estimates \hat{h}_k defined above in a manner to be described presently. The smoothed estimate can be shown to be a maximum likelihood estimate under certain restrictive circumstances (see [1] and [5]) but there is no need to pursue this fact here.

Thus let \hat{h} be a given estimate of h defined on $[0, T]$. We construct a new estimate \bar{h} based on \hat{h} as follows: Let \bar{h} coincide with \hat{h} until the first point, say t_0 , at which \hat{h} increases. Then redefine \bar{h} to be the average value of \hat{h} on the intervals to the left and right of t_0 . Repeat this process as often as is necessary until \bar{h} is decreasing on the interval $[0, t_0]$, and then proceed to the next point of increase of h .

For example, suppose \hat{h} has the values shown in Figure 1. Then Figure 2 shows the smoothed estimate \bar{h} .

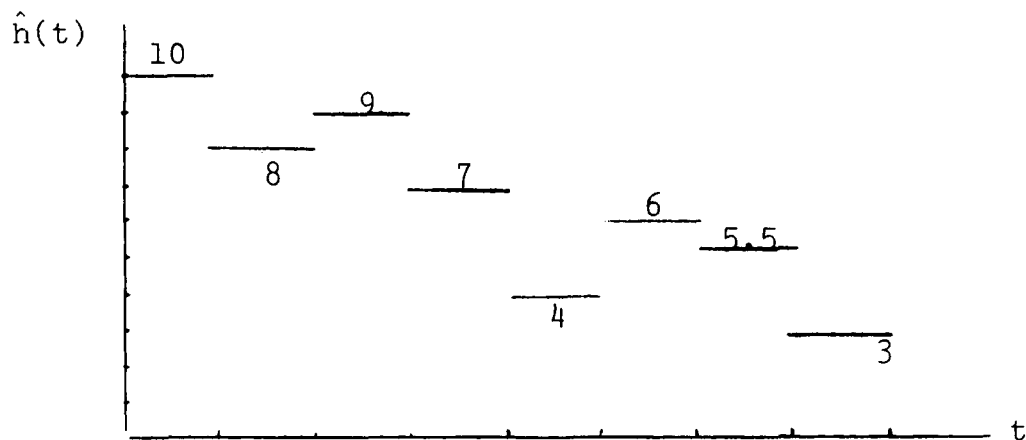


Figure 1

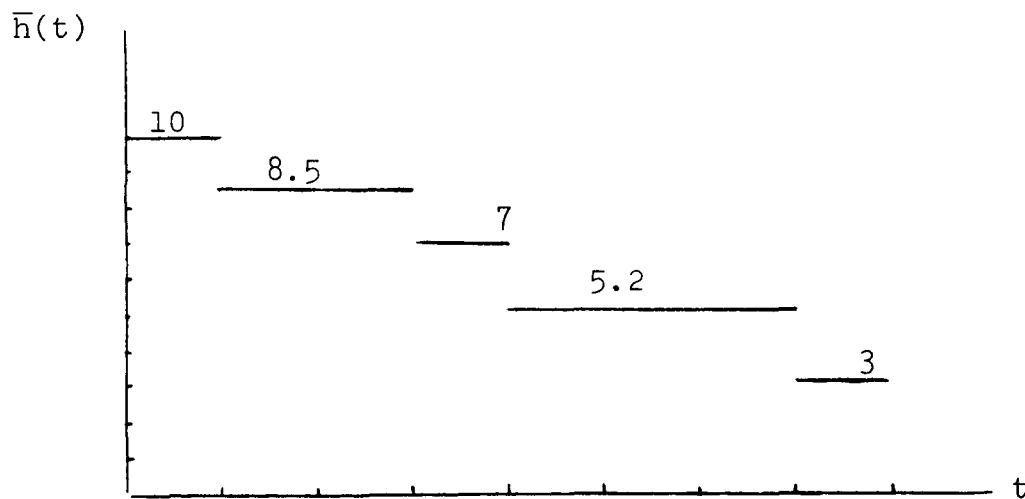


Figure 2

One of the principal virtues of the smoothed estimate is, of course, that it gives a better picture of the trend in the failure rate. As with all smoothing operations it irons out or dissipates chance fluctuations from the dominant trends. This particular smoothing will be used again in the next subsection.

3.2 Another Estimator for $h(t)$

The statistic to be described here is the one studied in [1] where it is mentioned as being a maximum likelihood estimator for a function closely related to our $h(t)$ but with certain restrictions on the X_n . These conditions are a little unnatural and almost impossible to verify but that should not deter us from trying out the estimator and seeing how it behaves in this more general problem.

Let S_n and X_n be as above, for $n=1, \dots, N$. Then let $h^*(t) = 1/X_{n+1}$ for $S_n < t \leq S_{n+1}$ and for $n=0, 1, \dots, N-1$.

The similarity of h^* to the estimates \hat{h}_k of 3.1 should be noted. Consider a kernel K like K_1 of 3.1, except that the quantity Δ is allowed to vary in such a way that the interval considered always contains one and only one failure point S_k and so that $\Delta = S_{k+1} - S_k = X_k$. Then

$$\begin{aligned}\hat{h}_k(t) &= \frac{1}{\Delta} \cdot (\# \text{ of failures in interval of length } \Delta \\ &\quad \text{about } t) \\ &= \frac{1}{\Delta} \cdot 1 = 1/X_k = h^*(t) \text{ if } S_k < t \leq S_{k+1}.\end{aligned}$$

The estimate h^* must be smoothed so that it becomes a decreasing function and the following averaging is recommended in [1]: If $1/X_k < 1/X_{k+1}$ so that h^* must be smoothed on the interval $(S_{k-1}, S_{k+1}]$ then let $h^* = 2/(X_k + X_{k+1})$ there.

A comparison of the estimates \hat{h}_k and h^* should really await some extensive testing but even at this vantage point some remarks are called for. A small scale trial of some fictional data seems to indicate that h^* gives a smoother more easily interpreted graph and would be preferable over h_k for that reason if this phenomena persists. There is also the fact that h^* is much easier to compute which may be an overriding consideration. It should also be mentioned that for visual enhancement, a piecewise linear graph is easier to look at than a step-function despite whatever mathematically optimal properties the latter has, and so one should not hesitate to "connect the dots" when attempting a visual evaluation.

4. STATISTICAL TESTS

In this section we assume that conditions (A), (B) and (C) are all fulfilled and comment on some statistical problems of obvious interest. In each case assume as given the failure data up to time T , that is the function $N(t)$, $0 \leq t \leq T$.

Consider first the problem of estimating μ . It seems clear intuitively that for the smoothed estimates \bar{h} and h^* , the values $\bar{h}(T)$ and $h^*(T)$ converge to $1/\mu$ almost surely, although no proof will be offered for this conjecture at this time. Thus $1/\bar{h}(T)$ or $1/h^*(T)$ seems to be a reasonable estimate for μ at time T . The speed of convergence will depend on how long the breaking-in process persists, and, in the case of \bar{h} , to what extent recent data get emphasized. Once more it must be emphasized that for the general problem considered here it is impossible to give any more precise information on the moments or distributions of these estimates.

Another problem of great interest is of a decision theoretic nature: has, as of time T , the breaking-in process ended? If the function $h(t)$ continues to decrease, for all t , to its limit $1/\mu$, e.g., $h(t) = 1/\mu + e^{-t}$, then of course the breaking-in never really ends so this question must be phrased a little more delicately. It can be done as follows. Let $\epsilon > 0$ be chosen according to some a priori standards. Then we agree to say that breaking-in has ended at the point $t^* = t^*(\epsilon)$ at which $h(t)$ first comes within ϵ of its eventual limit $1/\mu$. Thus $t^*(\epsilon) = \min_t \{h(t) - 1/\mu \leq \epsilon\}$.

Suppose h^* is the estimate of h being used. Then an obvious estimate for $t^*(\epsilon)$ would be

$$\underline{t}^* = \underline{t}^*(\epsilon, T) = \min_t \{h^*(t) - h^*(T) \leq \epsilon\}.$$

Thus \underline{t}^* is the smallest value of t such that the variation of h^* between t and the present T is not more than ϵ .

When $T < \underline{t}^*$, that is, breaking-in is still in progress then \underline{t}^* should be fairly close to T . Thus when \underline{t}^* stays close to T one should conclude that break-in has not yet ended. However, when T becomes bigger than \underline{t}^* and as T continues to grow, then \underline{t}^* begins to recede and becomes remote from T and when this phenomena is noted one should conclude that break-in is over.

5. CONCLUSIONS

Even with a bare minimum of assumptions about the distributions of times between failures some progress can be made in attempting to determine when a breaking-in process has terminated. The tests presented should be exercised with some simulated data in order to better understand how they perform and how they compare to one another. This is especially necessary since the generality of the problem has so far prevented any exact evaluation of the distributions or moments of these tests.



B. J. McCabe

1033-BJM-jr

Attachment
References

BELLCOMM. INC.

REFERENCES

1. Barlow, R., Proschan, F., and Scheuer, E., Maximum Likelihood Estimation and Conservative Confidence Interval Procedures In Reliability Growth and Debugging Problems, Rand Corporation Memorandum RM-4749-NASA, 1966.
2. Cox, D. R., Renewal Theory, Methuen & Co., 1962.
3. Kawata, T., A Renewal Theorem, Journal of the Mathematical Society of Japan, Vol. 8, 1956, pp. 118-126.
4. Loève, M., Probability Theory, Van Nostrand, 1955.
5. Marshall, A., and Proschan, F., Maximum Likelihood Estimation for Distributions with Monotone Failure Rate, Annals of Math. Stat., 1965, pp. 69-77.
6. Parzen, E., On Estimation of a Probability Density Function and Mode, Annals of Math. Stat., 1962, pp. 1065-1076.
7. Smith, W., On the Elementary Renewal Theorem for Non-Identically Distributed Variables, Pacific Journal of Mathematics, 1964, pp. 673-699.